

Exhibit D

Part 4

Correlation

290-NBF Document 108-12 Filed 04/16/1

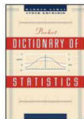
Claim Term	CMU's Construction	Marvell's Construction
correlation	the degree to which two more items (here, noise in signal samples) show a tendency to vary together.	the expected (mean) value of the product of two random variables (e.g., $E[r_i r_j]$, where r_i and r_j are signal samples at time i and time j , respectively).
'839 Patent Claims 11, 16, 19, 23 '180 Patent Claim 6	CMU Brf. at 19-20	Marvell Brf. at 17-21

- The Dispute:
 - Should “correlation” be accorded its ordinary meaning in engineering and statistics (Marvell) or its lay meaning (CMU)?

- Marvell's constructions are expressed as a ***single mathematical formula for measuring the extent to which two things are correlated,*** it is ***not*** a definition of "correlation"

- Marvell cites to the *Pocket Dictionary of Statistics* for the technical terms: "covariance," "mean," "variance," "matrix," and "covariance matrix"

Marvell's Opening Claim Construction Brief at pgs. 5, 7, 9, 20, 22, and 26.

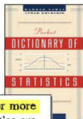


- Marvell ignores the definition of "correlation" found in this same reference

- Marvell's constructions are expressed as a *single mathematical formula for measuring the extent to which two things are correlated*, it is not a definition of "correlation"

- The *Pocket Dictionary of Statistics* defines "correlation" as:

correlation— A general term denoting association or relationship between two or more variables. More generally, it is the extent or degree to which two or more quantities are associated or related. It is measured by an index called **correlation coefficient**. See also *intraclass correlation*, *Kendall's rank correlation*, *Spearman's rank correlation*.



McElhinny Declaration, 1/27/10, Ex. 10.

- Correlation-sensitive branch metrics calculated from noise covariance matrices

11. A method for detecting a sequence that exploits the correlation between adjacent signal samples for adaptively detecting a sequence of symbols stored on a high density magnetic recording device, comprising the steps of:

- performing a Viterbi-like sequence detection on a plurality of signal samples using a plurality of **correlation-sensitive branch metrics**;
- outputting a delayed decision on the recorded symbol;
- outputting a delayed signal sample;
- adaptively updating a plurality of noise covariance matrices in response to said delayed signal samples and said delayed decisions;
- recalculating said plurality of **correlation-sensitive branch metrics** from said noise covariance matrices using subsequent signal samples; and
- repeating steps (a)–(e) for every new signal sample.

See '839 Patent Claims 11, 16, 19, 23; '180 Patent Claim 6



Document 108-12

US 6,201,839 B1

$$y_i = u_i + n_i, \quad i = 1, 2, \dots, N$$

In the derivations of the branch metrics (10), (13) and (15), no assumptions were made on the exact Viterbi-type

algorithm used for the Viterbi-type algorithm, and the assumption of a decision feedback equalizer is usually an implementation of the recursive least squares algorithm. Alternatively, the adaptation may be made using the least mean squares algorithm.

The quantities n_i that are subtracted from the output of the delay circuit 54 are the target response values, or mean

$$M_i = \log \det \frac{C_i}{\det c_i} + N_i^T C_i^{-1} N_i - u_i^T c_i^{-1} u_i \quad (13)$$

response values or mean response values or response values of the Viterbi-type detector 30. A non-circuit 55 computes the sum of the outputs of the circuits 30 and 52.

As stated above, the covariance matrix is given by:

$$C_i = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{bmatrix} \quad (14)$$

Using standard techniques of signal processing, it can be shown that:

$$\frac{\partial \det C_i}{\partial \sigma_i^2} = 2 \cdot \sigma_i^{-2} \cdot C_i \quad (15)$$

This ratio of determinants is referred to as α_i^T , i.e.:

$$\alpha_i^T = \frac{\partial \det C_i}{\partial \sigma_i^2} = 2 \cdot \sigma_i^{-2} \cdot C_i \quad (16)$$

It can be shown by using standard techniques of signal processing that the sum of the last two terms of (13), i.e. the output of the circuit 52, is:

$$L_i = \alpha_i^T C_i^{-1} N_i - u_i^T c_i^{-1} u_i \quad (17)$$

$$= \frac{\alpha_i^T N_i^T}{\sigma_i^2} \quad (18)$$

Where the vector α_i is (1,1)-dimensional and is given by:

$$\alpha_i^T = (1, 1) \quad \text{with} \quad \text{with} \quad \dots \quad \text{with} \quad (1, 1)^T \quad (19)$$

$$\alpha_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (20)$$

Equations (17), (19) and (18) (the circuit 52) can be implemented as a tapped-delay line as illustrated in FIG. 30. The circuit 52 has 1, delay circuit 54. The tapped-delay line implementation shown in FIG. 30 and 31 is also referred to as a moving-average, feed-forward, or finite-impulse response filter. The circuit 52 can be implemented using any type of filter as appropriate.

Algorithm

Computing the branch metrics in (10) or (13) requires knowledge of the signal statistics. These statistics are the mean signal values u_i in (12) as well as the covariance matrices C_i in (13). In magnetic recording systems, these statistics will generally vary from track to track. For example, the statistics that apply to a track at a certain radius will differ from those for another track at a different radius due to different line at track velocities at those radii. Also, the signal and noise statistics will be different if a head is flying slightly off-track or if it is flying directly over the track. The head skew angle is another factor that contributes to different statistics from track to track. These factors suggest that the system that implements the metric in (13) needs to be flexible to these changes. Storing the statistics for each track separately is very difficult because of the memory span required to accomplish this. A reasonable alternative is to use adaptive filtering techniques to track the needed statistics.

Tracking the mean signal values u_i is generally done so that these values fall on pre-specified targets. An adaptive filter and equalizer is employed to force the signal sample values to their targets. This is certainly the case with partial response targets used in algorithms like PR4, EPR4, or EPR4 where the target is pre-specified to one of the class-4 partial responses. For example, in a PR4 system, the signal samples, if there is no noise in the system, fall on one of the three target values 1, 0, or -1. Typically this is done with an LMS-class (least mean-squares) algorithm that ensures that the mean of the signal samples is close to these target values. In decision feedback equalization (DFE) based detectors or hybrids between feed delay line search and DFE, such as FDTs/DF or MDFE, the target response need not be pre-specified. Instead, the target values are chosen on-the-fly by simultaneously updating the coefficients of the front-end and feedback equalizers with an LMS-type algorithm.

When there are severe non-linearities in the system (also referred to as non-linear distortion or non-linear ISI), a linear equalizer will generally not be able to place the signal samples right on target. Instead, the means of the signal samples will fall at a different value. For example, in a PR4 system, the response to a sequence of written symbols $\dots, -1, 0, 1, 0, \dots$ might result in mean sample target values $\dots, 0, 1, 0, \dots$, while a sequence of written symbols $\dots, 1, -1, 0, \dots$ might result in a sequence of mean sample

Specification: Uses Mathematical Terminology

Euclidian branch metric. In the simplest case, the noise samples are realizations of independent identically distributed Gaussian **random variables** with **zero mean** and **variance σ^2** . This is a white Gaussian noise assumption. This implies that the **correlation distance** is $L=0$ and that the noise pdf's have the same form for all noise samples. The total ISI

'839 Patent 5:58-64

- random variables
- mean
- variance
- correlation distance
- covariance matrix
- expected values
- correlation-sensitive metric

The $(L+1) \times (L+1)$ matrix C_i is the **covariance matrix** of the data samples $r_i, r_{i+1}, \dots, r_{i+L}$, when a sequence of symbols $a_{i-KL}, \dots, a_{i+L+KL}$ is written. The matrix c_i in the denominator of (11) is the $L \times L$ lower principal submatrix of $C_i = [c_{ij}]$. The $(L+1)$ -dimensional vector \underline{N}_i is the vector of differences between the observed samples and their **expected values** when the sequence of symbols $a_{i-KL}, \dots, a_{i+L+KL}$ is written, i.e.:

$$\underline{N} = [(r_i - m_i)(r_{i+1} - m_{i+1}) \dots (r_{i+L} - m_{i+L})]^T \quad (12)$$

The vector \underline{n}_i collects the last L elements of \underline{N}_i , $\underline{n}_i = [(r_{i+1} - m_{i+1}) \dots (r_{i+L} - m_{i+L})]^T$.

With this notation, the general **correlation-sensitive metric** is:

$$M_i = \log \det \frac{C_i}{\det c_i} + \underline{N}_i^T C_i^{-1} \underline{N}_i - \underline{n}_i^T c_i^{-1} \underline{n}_i \quad (13)$$

'839 Patent 6:56-7:4



1. A method of determining branch metric values for branches of a trellis for a Viterbi-like detector, comprising: selecting a branch metric function for each of the branches at a certain time index; and applying each of said selected functions to a plurality of signal samples to determine the metric value corresponding to the branch for which the applied branch

metric function was selected, wherein each sample corresponds to a different sampling time instant.

2. The method of claim 1 further comprising the step of receiving said signal samples, said signal samples having signal-dependent noise, correlated noise, or both signal-dependent and correlated noise associated therewith.

3. The method of claim 1 wherein said branch metric functions for each of the branches are selected from a set of signal-dependent branch metric functions.

4. A method of determining branch metric values for branches of a trellis for a Viterbi-like detector, comprising: selecting a branch metric function for each of the branches at a certain time index from a set of signal-dependent branch metric functions; and applying each of said selected functions to a plurality of signal samples to determine the metric value corresponding to the branch for which the applied branch metric function was selected, wherein each sample corresponds to a different sampling time instant.

selecting a branch metric function for each of the branches at a certain time index; and applying each of said selected functions to a plurality of signal samples to determine the metric value corresponding to the branch for which the applied branch

metric function was selected, wherein each sample corresponds to a different sampling time instant; and applying each of said selected functions to a plurality of signal samples to determine the metric value corresponding to the branch for which the applied branch

10. A method of generating a branch weight for branches of a trellis for a Viterbi-like detector, wherein the detector is used in a system having Gaussian noise, comprising: selecting a plurality of signal samples, wherein each sample corresponds to a different sampling time instant;

calculating a first value representing a logarithm of a quotient of a determinant of a trellis branch dependent covariance matrix of said signal samples and a determinant of a trellis branch dependent covariance matrix of a subset of said signal samples;

calculating a second value representing a quadratic of said signal samples less a plurality of target values normalized by a trellis branch dependent covariance of said signal samples;

calculating a third value representing a quadratic of a subset of said signal samples less a plurality of channel target values normalized by a trellis branch dependent covariance of said subset of signal samples;

calculating the branch weight from said first, second, and third values; and

outputting said branch weight.

comprising:

- a Viterbi-like detector circuit, said Viterbi-like detector circuit for producing a plurality of delayed decisions and a plurality of delayed signal samples from a plurality of signal samples;
- a noise statistics tracker circuit responsive to said Viterbi-like detector circuit for updating a plurality of noise covariance matrices in response to said delayed decisions and said delayed signal samples; and
- a correlation-sensitive metric computation update circuit responsive to said noise statistics tracker circuit for calculating a plurality of correlation-sensitive branch metrics from said noise covariance matrices, said correlation-sensitive metric computation update circuit outputting a plurality of branch metrics to said Viterbi-like detector circuit.

24. A system for receiving, processing, and outputting a plurality of data, comprising:

- a write signal processing circuit for processing a plurality of data from a data source;
- a write control circuit;
- a write head responsive to said write control circuit for receiving a plurality of signals from said write signal processing circuit, said write head for writing said signals to the recording medium;
- a read control circuit;
- a read head for reading said signals from the recording medium, said read head responsive to said read control circuit; and
- a detector circuit for detecting a plurality of data from said read signals, said detector circuit having a circuit for

Specification: Branch Metric Computation

- Correlation is used in the computation of the branch metric (13)

- $E[\hat{C}(\hat{a})] = E[\underline{N}_i \underline{N}_i^T]$ calculates the expected value of the product of signal samples



need for further mean corrections. The focus is shifted to tracking the noise covariance matrices needed in the computation of the branch metrics (13).

Assume that the sequence of samples $r_i, r_{i+1}, \dots, r_{i+L}$ is observed. Based on these and all other neighboring samples, after an appropriate delay of the Viterbi trellis, a decision is made that the most likely estimate for the sequence of symbols $a_{i-K_f}, \dots, a_{i+L+K_f}$ is $\hat{a}_{i-K_f}, \dots, \hat{a}_{i+L+K_f}$. Here L is the noise correlation length and $K=K_f+K_r+1$ is the ISI length. Let the current estimate for the $(L+1) \times (L+1)$ covariance matrix corresponding to the sequence of symbols $\hat{a}_{i-K_f}, \dots, \hat{a}_{i+L+K_f}$ be $\hat{C}(\hat{a}_{i-K_f}, \dots, \hat{a}_{i+L+K_f})$.

This symbol is abbreviated with the shorter notation, $\hat{C}(\hat{a})$. If the estimate is unbiased, the expected value of the estimate is:

$$E\hat{C}(\hat{a}) = E[\underline{N}_i \underline{N}_i^T] \quad (21)$$

where \underline{N}_i is the vector of differences between the observed samples and their expected values, as defined in (12).

Prosecution History: Confirms Marvel's Construction

- Patent Office rejected CMU's claims over Huszar

The Examiner rejected claims 11-22 as being anticipated by U.S. Patent No. 5,862,192 to Huszar et al. The Examiner stated that Huszar et al. "discloses a method for detecting a sequence that exploits the correlation between adjacent signal samples for

6/12/00 Amdt. at 8, '839 Patent File History (Marvel Exh. 22)

- CMU argued that correlation requires multiplying signal samples

Huszar et al. discloses branch metrics that are not correlation sensitive. Instead, the branch metrics of Huszar et al. are path metrics that have the form of (See Huszar et al., col. 8, equation 17):

$$J = \sum_{\text{from } j=-\infty \text{ to } n} M_i$$

where M_i is a branch metric of the form:

$$M_i = [r_i(0) - y_i(0)]^2 + [r_i(1) - y_i(1)]^2$$

Such a branch metric is not correlation sensitive, as claimed in independent claims 11, 16, and 19, which is evidenced by the fact that there is no term in the branch metric that corresponds to the correlation between $r_i(0)$ and $r_i(1)$, i.e. there is no term that involves multiplying $r_i(0)$ with $r_i(1)$. Thus, Huszar et al. does not disclose branch metrics that are correlation sensitive. Furthermore, Applicants submit that Huszar et al. does not disclose the use of noise covariance matrices. Because Huszar et al. does not disclose branch

Id. at 8-9

- Marvell's file history argument undercuts its proposed construction
 - Marvell ignores the fact that the referenced multiplication of signal samples does not result in measuring the extent of a correlation
 - Marvell also ignores the fact that the file history discussion says nothing about using an "expected value," the term at the heart of Marvell's proposed construction

MARVELL'S PROPOSED CONSTRUCTION

"Correlation" means "the expected (mean) value of the product of two random variables (e.g., $E[r_i r_j]$, where r_i and r_j are signal samples at time i and time j , respectively)."

'839 Patent, at 1:38-67; 4:4-18; 4:43-47; 5:59-67; 6:5-20; 6:36-43; 6:53-65; 9:24-40; 13:3-7; 13:38-50. '839 File History, March 10, 2000 Office Action and Response thereto.

Extrinsic Evidence: Technical Treatises

- Marvell's construction is identical to statistical meaning:
 - ▶ $E[XY]$ = the expected (mean) value of the product of two random variables X and Y

The second-order moment $m_{11} = E[XY]$ is called the *correlation* of X and Y . It is so important to later work that we give it the symbol R_{XY} .

Peebles, *Probability, Random Variables, and Random Signal Principles*, at 102 (1980) (Marvell Exh. 23)

X . In electrical engineering, it is customary to call the $j = 1$ $k = 1$ moment, $E[XY]$, the **correlation of X and Y** . If $E[XY] = 0$, then we say that **X and Y are orthogonal**.

Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, at 233 (1994) (Marvell Exh. 18)

- ▶ See also Proakis Decl. at ¶¶ 30-31.

“Correlation” means “the expected (mean) value of the product of two random variables (e.g., $E[r_i r_j]$, where r_i and r_j are signal samples at time i and time j , respectively).”

Correlation

Correlation-sensitive branch metric. In the most general case, the correlation length is $L > 0$. The leading and trailing ISI lengths are K_l and K_r , respectively. The noise is now considered to be both correlated and signal-dependent. Joint Gaussian noise pdfs are assumed. This assumption is well justified in magnetic recording because the experimental evidence shows that the dominant media noise modes have Gaussian-like histograms. The conditional pdfs do not factor out in this general case, so the general form for the pdf is:

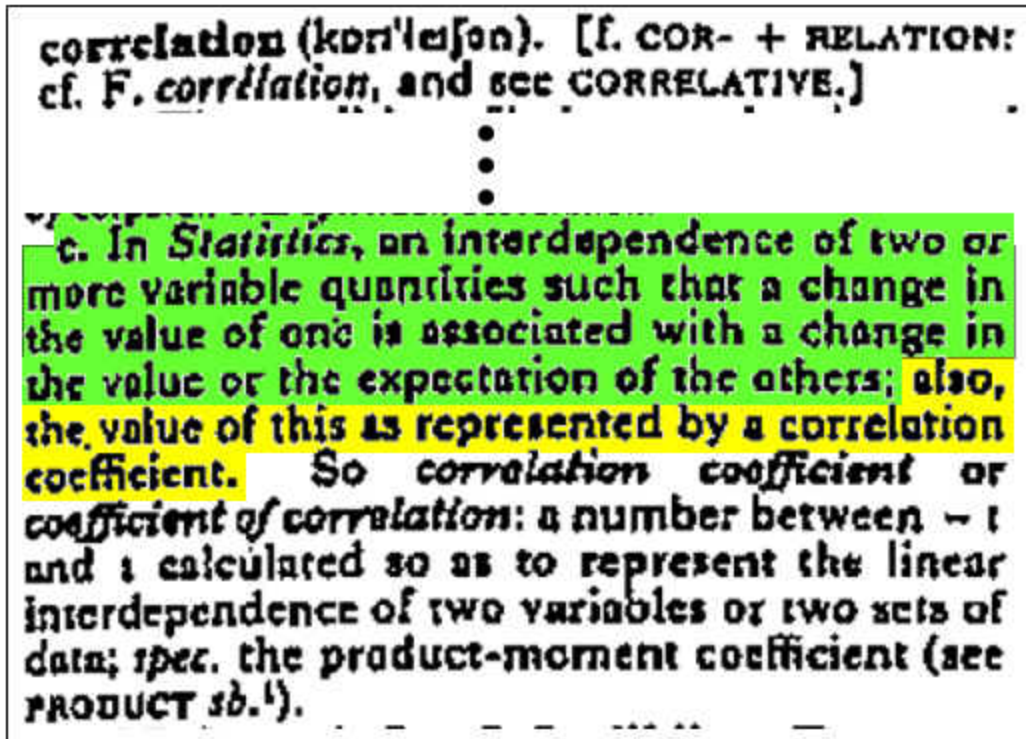
MARVELL'S PROPOSED CONSTRUCTION

“Correlation” means “the expected (mean) value of the product of two random variables (e.g., $E[r_i r_j]$, where r_i and r_j are signal samples at time i and time j , respectively).”

CMU's Reliance on General Dictionaries Fails

- CMU cites the Oxford English Dictionary

CMU Brf., at 21 n. 14



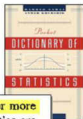
Compact Oxford English Dictionary (2d ed. 1987) (CMU Exh. 6)

- CMU truncated the definition that cited a “value”

- Marvell's constructions are expressed as a *single mathematical formula for measuring the extent to which two things are correlated*, it is not a definition of "correlation"

- The *Pocket Dictionary of Statistics* defines "correlation" as:

correlation— A general term denoting association or relationship between two or more variables. More generally, it is the extent or degree to which two or more quantities are associated or related. It is measured by an index called **correlation coefficient**. See also *intraclass correlation*, *Kendall's rank correlation*, *Spearman's rank correlation*.



McElhinny Declaration, 1/27/10, Ex. 10.